Projecting Global Datasets to Achieve Equal Areas

E. Lynn Usery, Michael P. Finn, John D. Cox, Thomas Beard, Sheila Ruhl, and Morgan Bearden

ABSTRACT: Scientists routinely accomplish global modeling in the raster domain, but recent research has indicated that the transformation of large areas through map projection equations leads to errors. This research attempts to gauge the extent of map projection and resampling effects on the tabulation of categorical areas by comparing the results of three datasets for seven common projections. The datasets, Global Land Cover, Holdridge Life Zones, and Global Vegetation, were compiled at resolutions of 30 arc-second, ½ degree, and 1 degree, respectively. These datasets were projected globally from spherical coordinates to plane representations. Results indicate significant problems in the implementation of global projection transformations in commercial software, as well as differences in areal accuracy across projections. The level of raster resolution directly affects the accuracy of areal tabulations, with higher resolution yielding higher accuracy. If the raster resolution is high enough for individual pixels to approximate points, the areal error tends to zero. The 30-arc-second cells appear to approximate this condition.

Introduction

Global modeling has become commonplace among scientists concerned with the environmental effects of human activities. Scientists routinely accomplish such modeling in the raster domain, using high resolutions for large parts of continents and low to high resolutions for the entire globe. Generally, low- to-moderate resolution datasets are 10-, 5-, or 1-degree cell size, and high-resolution datasets are 30-arc-second cells. Recent research has indicated that the transformation of such large areas through map projection equations and subsequent resampling leads to errors in statistical results tabulated from attributes of the transformed data (Steinwand et al., 1995; Usery and Seong 2001; Usery et al. 2002). A theoretical explanation of the transformation effects is given in Seong and Usery (2001), using some empirical data at the continental scale for Asia (Seong 1999). Kimerling (2002) developed a predictive model for the effects of pixel loss and duplication from transformations on equal-angular grids, and Seong et al. (2002) proposed the sinusoidal as one of the projections that reduce the problem. In Seong et al. (2002), errors were computed from a synthetic data matrix without the use of actual Earth surface datasets. Thus, only limited empirical work with real geographical data has been compiled tabulating the accuracy of total categorical areas in projected raster databases of global extent.

E. Lynn Usery, **Michael P. Finn**, **John D. Cox**, **Thomas Beard**, **Sheila Ruhl**, and **Morgan Bearden** work at the U.S. Geological Survey, Mid-Continent Mapping Center, 1400 Independence Road, Rolla, Missouri. Email: <usery@usgs.gov>.

This paper is an attempt to fill the gaps in our knowledge concerning the empirical results of projection transformation of regional and global raster datasets. The goal of this research is not to generate absolute accuracy in the final projected datasets; rather, the goal is to assess the errors introduced in the transformation and resampling process. We do not assume that data in equal-angular grids are the most accurate or the best starting point for any given analysis. As Mulcahy (2000) indicates, data in equal-angular grids may include distortions arising from transformations of original sources. We simply wish to determine if projection transformation from equal-angular grids in spherical coordinates to a plane system and the required resampling introduce significant error in categorical areas, and if there are error patterns associated with various resolutions or particular projection selections. The empirical data for our analysis could have been simply resampled versions of data classified from satellite images. However, since the data exist in geographic coordinates in equal-angular cells, they provide a

Any use of trade, product, or firm names in this paper is for descriptive purposes only and does not imply endorsement by the U.S. Government.

basis from which to empirically assess the errors introduced from projection transformation and resampling.

The next section of this paper details the empirical approach of comparing areas from spherical datasets in geographic coordinates with areas resulting after projection to a plane system. The third section provides descriptions of the datasets used and projections examined, and the fourth section discusses problems encountered in implementing the projection transformations in commercial software systems. The fifth section provides statistical analysis and graphical results. A final section draws conclusions from this work and provides some recommendations for the correct use of projection transformations and future research. Complete documentation of this research project and the datasets used are available at http://mciweb.er.usgs.gov/carto_research/projection/index.html.

Approach

Initially, we used available global datasets and commercial-off-the-shelf (COTS) software for projection transformation to determine the accuracy of areas. However, because of operational difficulties with COTS software, we include projection transformation results and analysis from an internal U.S. Geological Survey (USGS) software package for raster projection—Map Image or *mapimg*.

We have applied global area transformation and some regional transformations to a variety of datasets. In this paper, we focus our attention on categorical datasets of vegetation, life (ecological) zones, and land cover. The basic characteristics of these publicly available data are described in the next section and summarized in Table 1. These global datasets, and others, can be freely downloaded from the sites shown in the table. Since a primary concern was the tabulation of areas of various thematic classes of the Earth's surface for statistical analyses, we selected equal-area projections as a focus but included two compromise (nonequal-area, nonconformal) projections, the Robinson and the Van der Grinten II.[1] Table 2 lists the selected projections and their primary characteristics.

In addition to studying the effects of projection transformation, our approach examines resolution dependencies, since the three datasets occur at different pixel sizes—1 degree, ½ degree, and 30 arc-second. To achieve the approximate area of the Earth's surface in a plane system, we chose resolutions of square cells in the plane to maintain the closest pixel sizes to the spherical data representation. We computed these sizes from the equation for the length of the arc of a circle of latitude between two geographic longitudes (see, for example, Torge 1980):

$$\Delta L = \int_{\lambda_1}^{\lambda_2} N\cos\phi d\lambda = N\cos\phi(\lambda_2 - \lambda_1) \tag{1}$$

where N is the radius of curvature in the prime vertical for a rotational ellipsoid. For a spherical Earth, N is equal to the radius of the sphere; therefore, N = R = 6,370,997.0 m.

Thus, for one cell of longitude, $d\lambda$, calculated in radians, at the Equator ($\phi = 0°$):

$$\Delta L = R\cos 0° \lambda \tag{2}$$

Equation (2) yields cell sizes of 111.2 km, 55.6 km, and 926.6 m for the 1-degree, ½-degree, and 30-arc-second cell sizes, respectively.

To determine the accuracy with which projection transformations and resampling retain areal values, we first determined the area of each raster cell in spherical coordinates. From calculus, and particularly Usery and Seong (2001), the area of a surface of revolution, A(S), is:

$$A(S) = 2\pi\int_a^b x ds = 2\pi\int_a^b f(x)\sqrt{1 + f'(x)}dx \tag{3}$$

where, for a circle of revolution centered at (0,0):

$$f(x) = \sqrt{R^2 - x^2} \tag{4}$$

$$f'(x) = \frac{-x}{\sqrt{R^2 - x^2}} \tag{5}$$

Applying Equations 3 to 5 to the spherical Earth, with the same R used above, and integrating in bands of latitude, where a band corresponds to the latitudinal height of a pixel for a full 360°:

$$a = R^2 \sin\phi_1 \tag{6}$$

$$b = R^2 \sin\phi_2 \tag{7}$$

where:

$$\phi_1 = \text{latitude} \tag{8}$$

$$\phi_2 = \text{latitude}_2 = \phi_1 + \Delta\phi \tag{9}$$

These equations can be used to calculate the surface area A(S) for a band of latitude ($\Delta\phi$). To

[1] Descriptions and classifications of map projections can be found in Snyder (1987), Pearson II (1990), and Bugayevskiy and Snyder (1995).

Dataset	Resolution	Source	File Size (approx.)	Web Site
Land Cover	30 "	AVHRR	1 Gb	http://edcdaac.usgs.gov/glcc/globdoc2_0.html#geom
Life Zones	0.5 º	UNEP/GRID	1 Mb	http://www.grid.unep.ch/data/grid/gnv5.html
Vegetation	1.0 º	NCAR/Scientific Computing Division	128 Kb	http://dss.ucar.edu/datasets/ds765.0/

Table 1. Global datasets used to assess map projection effects.

Projection	Classification	Property Preserved	Lines (Points) of True Scale	Use
Geographic		All		All
Goode	Pseudocylindrical	Area	Equator and parallels at 40º 44′ N&S	Global
Hammer	Pseudoazimuthal	Area		Global
Mollweide	Pseudocyclindrical	Area	Parallels at 40º 44′ N and S	Global
Sinusoidal	Pseudocyclindrical	Area	Central meridian, all parallels	Global,South America, Africa
Wagner IV	Pseudocyclindrical	Area	Equator	Global
Robinson	Pseudocyclindrical	Compromise	Equator	Global
Van der Grinten	Miscellaneous	Compromise	Equator	Global

Table 2. Projections examined.

determine the area of one pixel (square cell) in this band, we calculate:

$$A(S)_{cell} = \frac{A(S)_{band}}{\frac{360^{o}}{\Delta\lambda^{(o)}}} \qquad (10)$$

where:

$$\Delta\lambda = \Delta\phi \quad \text{(in degrees)} \qquad (11)$$

To illustrate, for a 1º latitude band ($\Delta\phi$), from 0º to 1º, the area of a single cell $A(S)_{cell}$ = 1.2363371878 x 10^{10}m^2. Table 3 presents some sample values of single-cell areas for 30-arc-second, ½-degree, and 1-degree cells at various latitudes. To verify the cell sizes calculated from Equation (2), we used these area values for each resolution cell at the Equator in Table 3, determined their square roots, and obtained matching sizes of 111.2 km, 55.6 km, and 926.6 m for the 1-degree, ½-degree, and 30-arc-second cell sizes, respectively.

From these calculated values, we created a separate data layer of a geographic information system (GIS) in which the pixel value of the cell is the area of that cell. For a specific thematic category, we computed the total area of the category by adding all cell areas occupied by that category. This provided the true geographic area on the spherical surface for each category and became the standard for testing the projected areas.

Latitude	30 "	0.5 º	1 º
0	858,631	3,091,035,692	12,363,671,878
15	829,390	2,982,220,448	11,970,315,668
30	743,628	2,670,171,821	10,761,202,175
45	607,188	2,176,155,408	8,818,730,582
60	429,370	1,533,837,609	6,275,272,108
75	222,291	786,991,318	3,304,173,896
90	63	13,487,417	107,896,706

Table 3. Areas of single pixels computed from spherical coordinates.

Dataset Descriptions

Global Land Cover Data

The global land cover data were obtained from the USGS Earth Resources Observation System (EROS) Data Center. The dataset consists of unsigned 8-bit values in a 30-arc-second grid. Coordinates are geographic latitude and longitude on a sphere of radius 6,370,997 m (Loveland et al. 2000). The dataset was developed as part of the National Aeronautics and Space Administration Earth Observing System Pathfinder Program and the International Geosphere-Biosphere Programme. The dataset is derived from 1-km Advanced Very High Resolution Radiometer (AVHRR) data spanning a 12-month period.

Global Life Zones Data

The Holdridge Life Zones dataset is from the International Institute for Applied Systems Analyses in Laxenburg, Austria (Leemans 1990). The dataset shows the Holdridge Life Zones of the World, a combination of climate and vegetation (ecological) types. The Holdridge Life Zones data file consists of a ½-degree grid of data for the Earth with 8-bits per pixel.

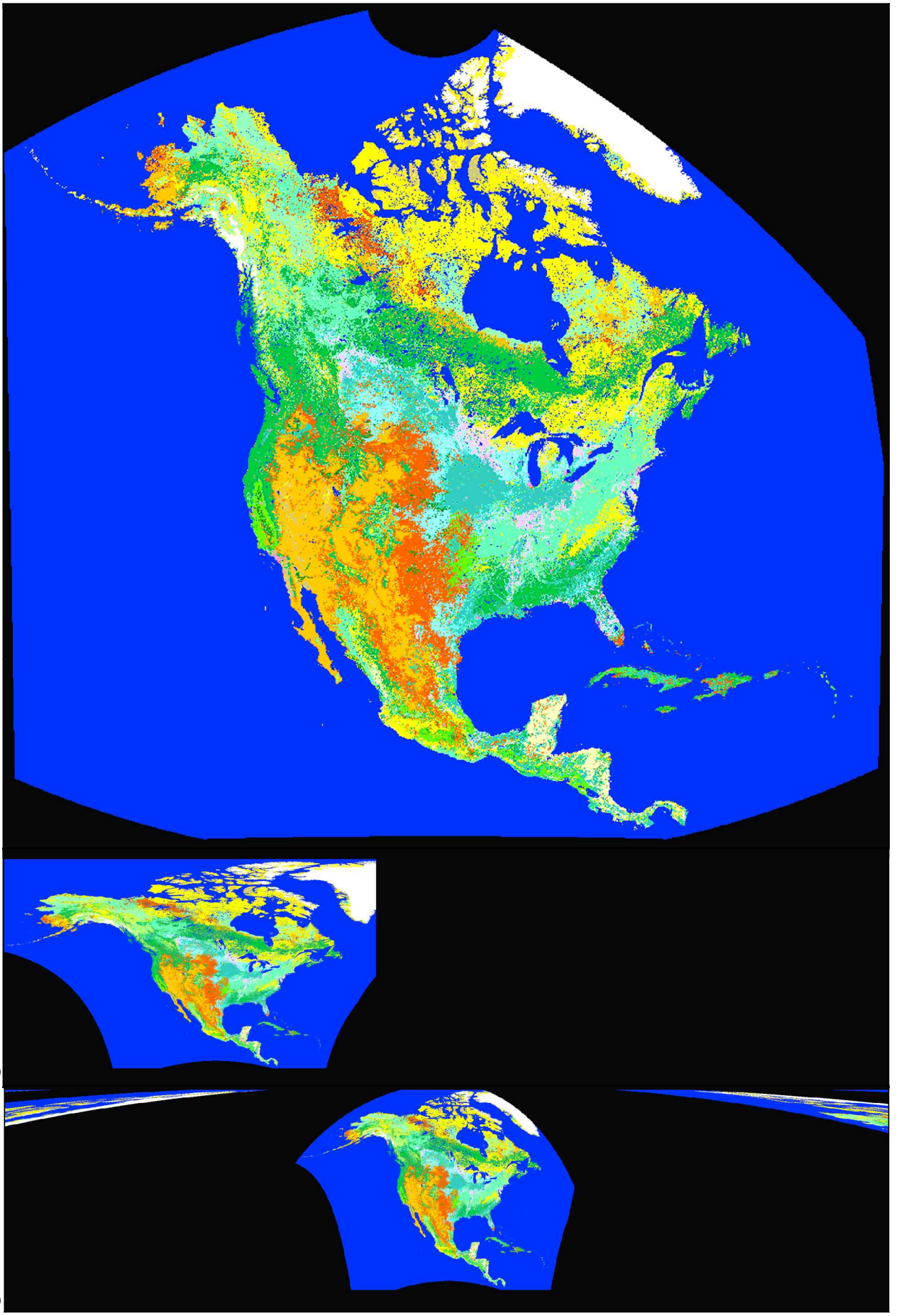

Figure 1. North American land cover in (a) Lambert Azimuthal, (b) inversely projected to geographic coordinates, and (c) forward projected to Mollweide.

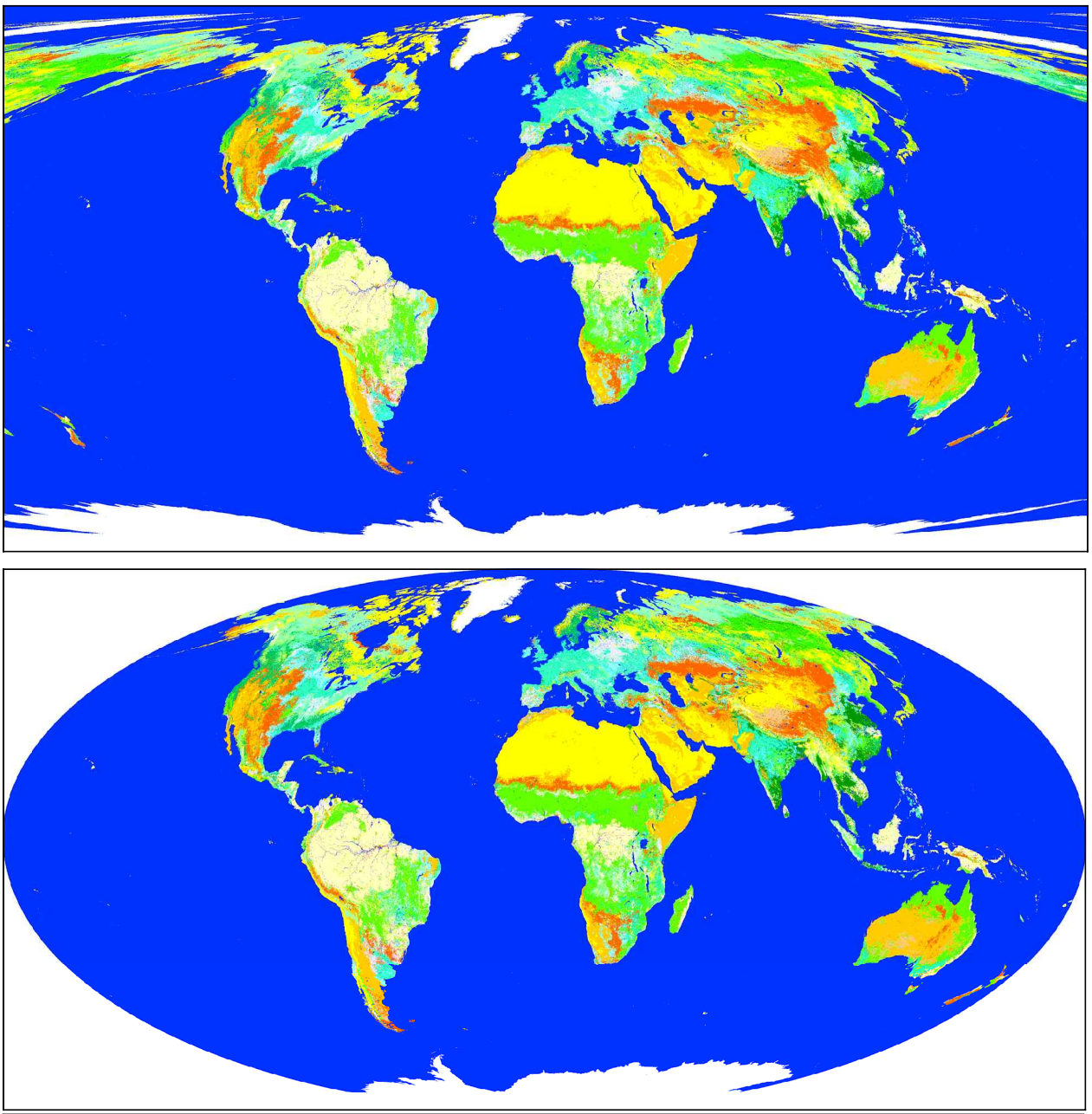

Figure 2. Projection of global land cover data to the Mollweide projection in (a) COTS package and (b) the USGS *mapimg* software.

Global Vegetation Data

This dataset of global vegetation was compiled on a 1-degree grid (Matthews 1983). It is archived at the National Center for Atmospheric Research and represents natural vegetation based on the United Nations Educational, Scientific and Cultural Organization classification system.

Results and Discussion

Initially, we assumed that the projection of raster data with commercial GIS software would be a simple process of setting the appropriate parameters and allowing the software to compute a new raster grid in our chosen projection. We quickly found numerous problems occur with this simple view. First, some software packages do not use exact projection equations and therefore cannot be used for global data. Polynomial approximations work well for small areas of the Earth's surface, but when the projected area becomes large (about 400 x 400 km), these approximations cannot account for the distortion resulting from mapping a three-dimensional curved surface into a two-dimensional plane.

Exact projection equations must be used, followed by a resampling procedure appropriate for the data type in use.

The second problem encountered was that even when exact projection equations are used, the software often results in errors and sometimes "crashes" for specific projections, at specific resampling resolutions, and for specific singularities, such as the poles. Interestingly, the underlying base package for many commercial implementations is the USGS's General Cartographic Transformation Package (GCTP) designed for point-to-point transformations of vector data. Point-to-point transformation is the common approach because projection is a systematic transformation of spherical coordinates (ϕ,λ) to a plane coordinate (X, Y) representation. Mathematically,

$$X = f_1(\phi,\lambda), \qquad Y = f_2(\phi,\lambda) \tag{12}$$

defines a point transformation where X and Y are coordinates in the plane that are functions (f_1 and f_2) of ϕ and λ, latitude and longitude on the sphere.

Direct implementation of these equations with appropriate functions yields the variety of projections available in COTS, but implementation with data other than points, for example lines or grid cells, require specific adaptation of the equations or prior preparation of the data to allow the point-to-point transformation to succeed. For example, a line from New York to London representing the shortest distance between the two locations (a great circle route), shown in a spherical dataset as two points (ϕ_1, λ_1) and (ϕ_2, λ_2), will not transform to the desired straight line in the plane, i.e., the shortest distance, without the addition of many points between the two. The exception to this statement is when the projection transformation is defined specifically to transform great circles into straight lines, as in the gnomonic projection (Snyder 1987, p. 167):

$$X = R \cot \phi\lambda \sin (\lambda–\lambda_0) \tag{13}$$

$$Y = -R \cot \phi \cos (\lambda- \lambda_0) \tag{14}$$

where:

R = the radius of the sphere (Earth);

ϕ = latitude;

λ = longitude; and

λ_0 = the longitude of the central meridian.

Thus, the use of a point-to-point transformation for the map projection of lines requires specific consideration of the projection in question and the data to be projected.

Similarly, a projection of an equal-angular cell in a raster matrix cannot be accomplished directly as a point transformation unless the resolution is sufficient for the cell itself to be approximated as a point. Alternatively, the algorithm for the projection of cell data can account for resolution by treating each cell as a polygon bounded by lines and ensuring that the lines on the sphere are transformed correctly to straight lines bounding the cells in the plane.

Additional constraints apply to global data in either raster or vector formats. These include the appropriate handling of singularities for specific projection equations. In many projections, the pole becomes a line of pixels and thus distorts the data. Also, a continent, such as North America, may extend into the polar area, and thus conversion from geographic coordinates to a raster matrix may include the extension of the entire dataset to show the extent at the pole. A masking routine that bounds the output raster data to the continental boundaries or a correlating transformation to ensure the accuracy, as described below, is needed to properly subset the output matrix.

It appears that vendors used the GCTP package or similar point transformations without adaptation to the specific characteristics of raster data. For example, Figure 1a shows land cover data for North America in the Lambert Azimuthal Equal Area projection, and Figure 1b shows the result of inverse projection to geographic coordinates. Note the extension of the raster file, in Figure 1b, where the software has filled pixels with zero values to include the entire Northern Hemisphere (360°) in the output projection. This inclusion increases the file size by a factor of almost 6 (82 Mb to 455 Mb) and provides no useful information. Since reprojection from one projection to another (for example, Lambert Azimuthal to Mollweide) requires inverse projection as a part of the exact transformation process, reprojection creates erroneous data in an output projection. An example of this is shown in Figure 1c, in which the Mollweide projection was created as a reprojection from the Lambert Azimuthal in Figure 1a. A simple adaptation of the software for raster data to correctly transform only the required data is described below.

Commonly, transformations of raster data including global projections are computed in an inverse fashion. The derived output raster matrix is created, and the transformation mapping back into the input space is defined. This process facilitates resampling because the output cell is mapped to surrounding input cells and the output cell value can be interpolated directly. Unfortunately, with global data, such a mapping can result in an erroneous position and replicate data or create

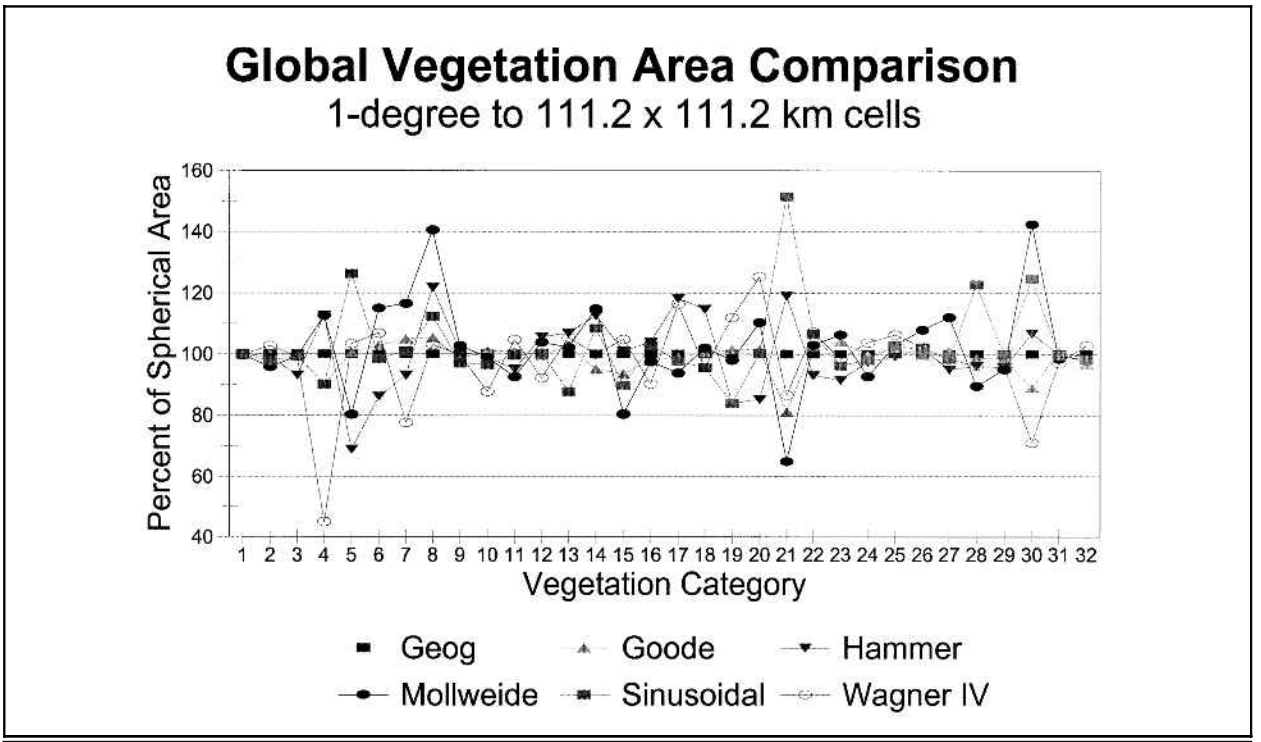

Figure 3. Comparison of projected areas of global vegetation categories.

pixels not in the original space. The solution is to perform a corresponding forward transformation to ensure that the same location results from both methods. This is the approach used in the USGS system, *mapimg* (Steinwand, personal communication 2003).

Other problems include projections that will not execute on global data. For instance, some projection selections work only at specific resampling resolutions. In one case, we could project global land cover data to the sinusoidal projection with a 1-, 16-, or 25-km output pixel size, but not with 4- or 8-km pixels. Another problem is that computation times may be lengthy, even on high-end personal computers. Global land cover (approximately a 1-Gb file) took 100 to 200 hours to be projected for some specific selections on a computer with dual 500-MHz processors.

A final problem is best illustrated with the graphics in Figure 2. Figure 2a shows the result of using commercial software to project global land cover data from geographic coordinates to a Mollweide projection. Note the replicated areas of Alaska and Russia at both sides of the map. Figure 2b shows the same projection operation but implemented correctly using *mapimg*. Although *mapimg* is also based on GCTP and point-to-point transformation, it uses special code to check the transformation in the forward direction after the inverse transformation. The remaining projections in Table 2 suffer similar distortion results from projection in commercial GIS software packages. We experimented with various COTS software packages with similar results.

Statistical Analysis and Comparison

Obviously, any comparison of statistical results for global raster datasets would be meaningless with outputs such as those shown above from commercial packages that generate replicated or distorted areas, as in Figures 1 and 2. However, it is relevant to examine results from *mapimg* and compare the various projections and the various resampling resolutions. For this purpose, we rely on the global thematic datasets (vegetation, life zones, and land cover), and we use the global projections listed in Table 2.

For all datasets, the process to determine the accuracy requires calculating the total area of each category in the projected space. Because all pixels in the projected raster space are of equal size, the computation was completed by determining the number of cells per category and multiplying by the pixel size. We then compared the areas of each category with the area resulting from the numerical integration of the data in spherical coordinates by calculating percentages for each category. Those percentages are the basis of the tables and figures in the following discussion.

	Goode	Hammer	Mollweide	Sinusoidal	Wagner IV	Robinson	Van der Grinten II
Minimum	81.16	68.90	64.93	83.95	45.05	90.10	50.68
Maximum	105.40	121.80	142.36	151.51	125.30	247.99	1354.21
Average	98.97	99.86	102.25	102.64	98.04	135.03	231.31

Table 4. Summary of percentage areas of vegetation categories by projection.

Global Vegetation Data Analysis — 1-Degree Cells

Table 4 and Figure 3 provide a summary of vegetation area comparisons across projections generated with *mapimg* from data in geographic coordinates. The vegetation classification containing 32 categories was projected from 1-degree spherical cells to 111.2 x 111.2 km cells in plane space. We do not assume the vegetation classification to be accurate but, merely, that if we have a numerical value for the total area occupied by a category in geographic coordinates, that value should be replicated as the area for the same category in the projected data. For the projected vegetation data, Table 4 provides a summary for categories of vegetation of the minimum, maximum, and average percentages of the spherical areas computed for each projection. For example, for the sinusoidal projection, the maximum areal deviation is 151.51 percent of the corresponding category in geographic coordinates. This category (number 21 in Figure 3) is "Cold-deciduous subalpine/subpolar shrubland, cold-deciduous dwarf shrubland" and represents a vegetation type occurring in high latitudes with a small total area of global coverage. With such a small area, slight areal differences create large percentage differences. Similarly, on the sinusoidal projection the small total area of the "Temperate/subpolar evergreen rain forest" category (number 5) results in a difference of 123.31percent. The category "Ice" (number 32 in Figure 3) also occurs at high latitudes and would be expected to show large areal differences. It does in absolute numbers, but because a significant part of the Earth is covered with ice, the percentages do not show great differences.

Table 4 shows that for the equal-area projections the *mapimg* software transformations result in an accuracy range from 45.05 to 151.51 percent. The numbers in Table 4 indicate that the Goode maintains the smallest deviation range (81.16- to 105.40-percent accuracy) of the categorical areas from geographic coordinates. This observation is reflected in Figure 3, but it is evident that for some categories of vegetation, the Goode does not maintain the best accuracy. For example, for the category "Ice" (number 32), the Goode has the lowest accuracy (94.69 percent) of any of the equal-area projections. The other equal-area projections vary more significantly, with errors of approximately 30 to 40 percent for some categories. On the Wagner IV, "Subtropical evergreen rainforest" (category 4) is only 45.05 percent accurate. Data for the Robinson and Van der Grinten II projections are included in Table 4 to demonstrate the effects of the use of non-equal-area projections. These data include percentages as high as 248 and 1,354 percent, respectively. Because of these high percentages, these two projections were not included in Figure 4.

Global Life Zones Data Analysis – ½-Degree Cells

Table 5 and Figure 4 compare areas for projecting life zone data from ½-degree cells in geographic coordinates to raster cells of 55.6 x 55.6 km. The life zone data are categorized into 38 different life zones or climate/vegetation (ecological) types. The table shows the Hammer with the smallest average deviation (99.62 percent of the spherical areas), but again the Goode has the smallest range of deviation (from 97.17 minimum to 103.25 maximum). Note that the sinusoidal and the Goode (a combination of the sinusoidal and Mollweide) both maintain a small range of areal error that is reflected in the graph in Figure 4. In addition, in Figure 4, the Wagner IV shows only 82.53 percent of the spherical area for the "Subtropical desert" category (number 24), but for the same category the Hammer shows 107.92 percent. The

	Goode	Hammer	Mollweide	Sinusoidal	Wagner IV	Robinson	Van der Grinten II
Min	97.17	94.44	87.79	96.73	82.53	101.57	101.86
Max	103.25	107.92	112.72	103.25	107.29	186.84	561.09
Avg	99.50	99.62	99.08	99.44	99.45	131.57	194.48

Table 5. Summary of percentage areas of life zone categories by projection.

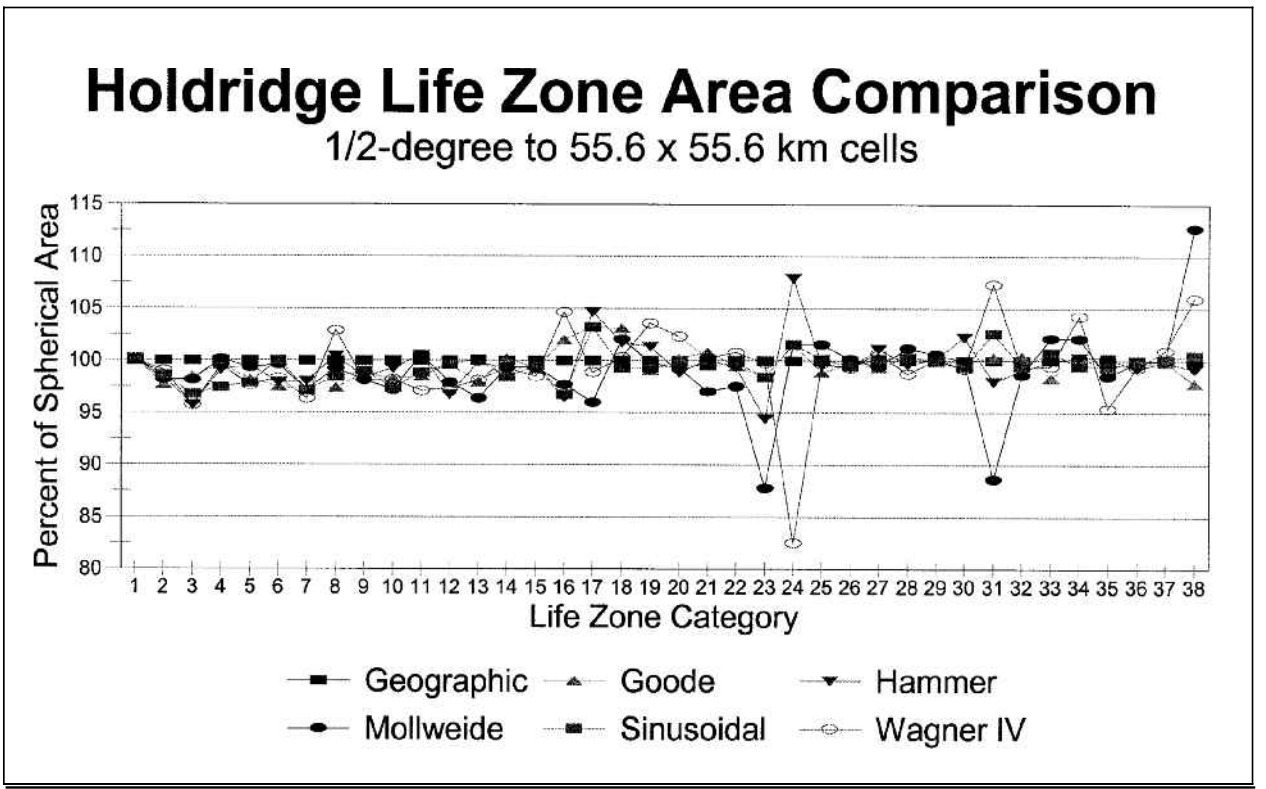

Figure 4. Comparison of projected areas of global life zones categories.

Mollweide shows 112.72 percent for the "Tropical Rainforest" category (number 38). The Robinson and Van der Grinten II again have higher errors, as expected.

It is significant to note that all five of the equal-area projections have increased accuracies with ½ degree cells over the 1-degree cells. However, for the Robinson and Van der Grinten II projections, the maximum errors are in the same range, 190 percent and 550 percent, respectively.

Global Land Cover Analysis – 30-Arc-second Cells

The same type of comparison of spherical-to-projected areas for global land cover is shown in Table 6 and Figure 5. It is important to note that the scale of the Y-axis only ranges over 0.8 percent in the graph of Figure 6. The land cover data consisting of 24 categories were projected from 30-arc-second spherical cells to 926.6 x 926.6 m cells in the plane space. In this case, the resolution of 30-arc-second appears high enough that the point transformations of the projection software do not introduce significant errors for the equal-area projections. For these projections (Figure 5), the minimum percentage error occurs for category number 23, "Bare ground tundra," on the Goode projection, with 99.59 percent of the area of the same category in spherical coordinates. This high latitude category, which occupies a small total geographic area, also scores the lowest percentage on the Mollweide (99.62 percent) and Wagner IV (99.60 percent). The maximum percentage among the equal-area projections is 100.27 for the same category on the Hammer projection. Figure 5 also shows that other categories with small total areas, for example, "Herbaceous wetland" (category number 17), show the greatest deviation**.**

All five equal-area projections achieve projection accuracy ranges within ±0.5 percent of the spherically computed areas. Note that there appears to be a systematic bias of approximately 0.2 percent for all the equal-area projections. This insignificant error can be explained by limitations resulting from nearest neighbor resampling and calculations performed in the discrete mathematical domain, such as the inclusion or exclusion of a pixel (area) along a border (line)—normally determined by whether the center of a pixel is "inside" the area, or not. A similar effect, a slightly larger error with 99.7-percent accuracy, is present in the global vegetation data, but the projection errors overwhelm the nearest neighbor resampling effect and the systematic error does not show in the illustration in Figure 3. In the life zone data, the resampling effect of pixels is almost balanced both outside and inside and the graph centers

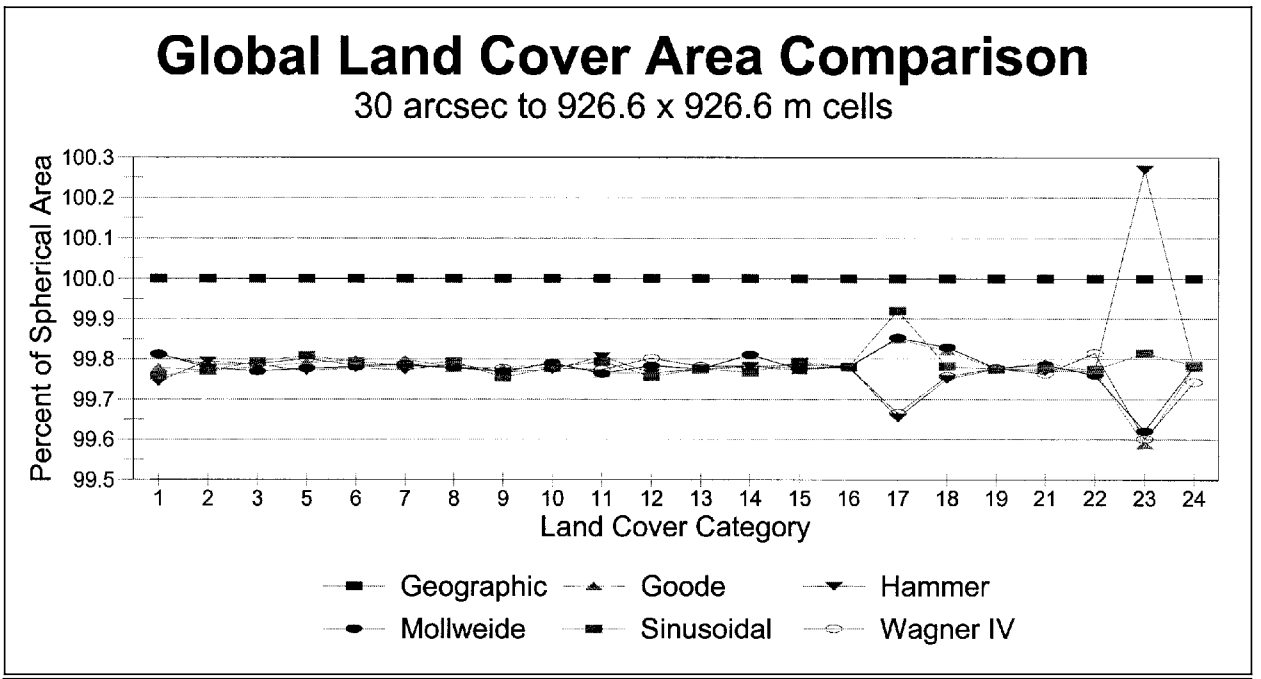

Figure 5. Comparison of projected areas of global land cover categories.

	Goode	Hammer	Mollweide	Sinusoidal	Wagner IV	Robinson	Van der Grinten II
Minimum	99.59	99.65	99.62	99.75	99.60	111.39	105.59
Maximum	99.85	100.27	99.85	99.92	99.81	250.43	1351.49
Average	91.46	91.48	91.46	91.47	91.45	132.47	265.69

Table 6. Summary of percentage areas of land cover categories by projection.

around 100.02 percent. The non-equal-area projections again show high errors—250 percent for the Robinson and 1,351 percent for the Van der Grinten II.

Conclusions

The problems of projecting global raster datasets with conventional GIS software are significant and alter the results of statistical area tabulation and percentages of categories. Various problems occur with such software, not the least of which is the lack of guidance for users in selecting projections appropriate for specific data and selecting appropriate parameters for the chosen projection. Online guides included with the software could resolve this problem, or a link to a Web system for map projection selection (USGS 2003a) could also provide appropriate guidance. Aside from the problems with implementation, and the extent of user knowledge and education, significant problems occur with maintaining categorical areas in global raster data. This areal distortion results directly from projection transformation and resampling. The results of these transformations and the extent of the inaccuracies are dependent on the input resolution of the dataset and the particular projection chosen. As resolution increases, the errors in categorical areas decrease.

Our analysis of global raster datasets of 1-degree vegetation, ½-degree life zone, and 30-arc-second land cover indicates that the equal-area projection of data in spherical coordinates with standard point-based transformations is only appropriate for high-resolution equal-angle grids. These results indicate that equal-area projections generate errors ranging from 18.84 to 54.95 percent (81.16- to 45.05-percent accuracy) for 1-degree cells, 2.83 to 17.47 percent (97.17- to 82.53-percent accuracy) for ½-degree cells, and less than ±0.5 percent for 30-arc-second cells. Non-equal-area projections in our testing of the Robinson and Van der Grinten II are inappropriate at all resolutions and produce maximum deviations of 250 and 1,354 percent, respectively.

Results indicate that conventional map projections based on point-to-point transformations are only appropriate for global raster datasets when

the resolution of the raster cells is high enough to approximate the point transformation. This analysis indicates that 30-arc-second cells are of sufficient resolution and result in errors of less than 0.5 percent in categorical areas.

Future work should fill the gaps and determine the exact cell sizes at which significant error is introduced. For example, if data are 10 arc-second or 5 arc-second, do the errors become zero? Additional work is also needed with data other than categorical datasets. For example, we have begun testing global elevation data from GTOPO30 (USGS 2003b). We know areal errors occur, but are elevations shifted in location? That is, are the peaks of mountains lost or moved in the transformation process? Since the data are numerical, the effects are different from the categorical datasets tested in this paper, and averaging resampling methods can be used. What types of errors are introduced by these processes? Finally, can we use the theoretical results from previous work (*e.g.*, Seong and Usery 2001; Kimerling 2002) to explain the empirical results? We have begun this assessment, but it will require significant research and testing since the empirical data, *i.e.*, categories, are distributed geographically, not regularly or in a systematic grid as were those used in the theoretical work.

ACKNOWLEDGMENT

The authors gratefully acknowledge the contributions of Daniel Steinwand, USGS EROS Data Center, who developed the *mapimg* projection software and helped in its use with the datasets described in this paper.

REFERENCES

Bugayevskiy, L.M., and J.P. Snyder. 1995. *Map projections: A reference manual.* London, U.K.: Taylor & Francis, Ltd.

Kimerling, A.J. 2002. Predicting data loss and duplication when resampling from equal-angle grids. *Cartography and Geographic Information Science* 29 (2) 111-26.

Leemans, Rik.1990.Global data sets collected and compiled by the Biosphere Project. Working Paper, International Institute for Applied Systems Analyses, Laxenburg, Austria.

Loveland, T.R., B.C. Reed, J.F. Brown, D. O. Ohlen, J. Zhu, L. Yang, and J.W. Merchant. 2000. Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR Data. *International Journal of Remote Sensing* 21(6/7):1303-30.

Matthews, E. 1983. Global vegetation and land use: new high-resolution data bases for climate studies. *Journal of Climatology and Applied Meteorology* 22: 474-87.

Mulcahy, K.A. 2000. Two new metrics for evaluating pixel-based change in data sets of global extent due to projection transformation. *Cartographica* 37(2): 1-11.

Pearson II, F. 1990. *Map projections: Theory and applications*. Boca Raton, Florida: CRC Press. 372 p.

Seong, J.C. 1999. *Multi-temporal, integrated global GIS database and land cover dynamics, Asia, 1982-1994*. Unpublished Ph.D. Dissertation, University of Georgia, Athens, Ga.

Seong, J.C., and E. L. Usery. 2001. Modeling raster representation accuracy using a scale factor model. *Photogrammetric Engineering and Remote Sensing* 67(10):1185-91.

Seong, J.C., K.A. Mulcahy, E. L. Usery. 2002. The sinusoidal projection: A new meaning for global image data, *The Professional Geographer* 54(2): 218-25.

Snyder, J.P. 1987. *Map projections—A working manual*. U.S. Geological Professional Paper 1395. U.S Government Printing Office, Washington, D.C. 383 p.

Steinwand, D.R. 2003. Personal conversation on Jan 10, 2003.

Steinwand, D.R., J.A. Hutchinson, and J.P. Snyder. 1995. Map projections for global and continental data sets and an analysis of pixel distortion caused by reprojection. *Photogrammetric Engineering and Remote* Sensing 61(12): 1487-97.

Torge, W. 1980. *Geodesy—An introduction*. Berlin, Germany: Walter de Gruyter.

Usery, E.L., M. Finn, and D. Scheidt. 2002. Projecting global raster databases. *Proceedings of Symposium on Geospatial Theory, Processing and Applications*, Ottawa, Canada.

Usery, E.L, and J.C. Seong. 2001. All equal-area map projections are created equal, but some are more equal than others. *Cartography and Geographic Information Science* 28(3): 183-93.

USGS (U.S. Geological Survey). 2003a. *Decision support system for map projections of small-scale data*. [http://isis.er.usgs.gov/research.dss_map.html, accessed 08 January 2003].

USGS (U.S. Geological Survey). 2003b. *Global 30-arc-secondond elevation data set*.[http://edc.usgs.gov/glis/hyper/guide/gtopo_30, accessed 08 January 2003].